

Melindan automaattiset korjaukset

Kansalliskirjaston asiantuntijaseminaari
25.11.2014
Ulla Ikäheimo

Taustaa

- Laatukeskustelu sähköpostilistoilla ja viimeksi Kirjastoverkkopäivien Melinda-työpajassa
 - MARC 21 – ja ISBD-konversiosta jääneet virheet
 - Formaatin muutokset, uudet kentät/osakentät, luettelointikäytäntöjen muutokset
 - Tietokantojen laajat aineistoajat, ammattikorkeakoulukirjastot, Helka, musiikkiaineistot, tulossa yleiset kirjastot
- > tuplatietueet ja tuplakentät, jotka vaikeuttavat tiedonhakua ja tietueiden käsittelyä
- > kehitetään automaattisia korjauksia

Tietueajoissa tapahtuu

- Kun tietueita ajetaan tietokantaan, merge-kentät (mm. 5-alkuiset) normalisoidaan = poistetaan välimerkit ja yhtenäistetään isot ja pienet kirjaimet
 - > Aleph säilyttää pidemmän kentän (välimerkillisen) tai yhtä pitkistä kentistä ensimmäisen ja toinen kenttä poistetaan
- Eri tavalla muotoillut fraasit huomautuksissa, eri kielet, eri osakentät, kirjoitusvirheet, yhteensidotut ym. tekevät edelleen tuplakenttiä

Suunnitellut kenttäkorjaukset

- Kiinteiden kenttien formaattikonversiovirheet
000/19 'r' -> tyhjä, 008/23 'r' -> ei koodattu
- Yhtenäistetty nimeke 240 -> 130, kun teoksella ei tekijää
- Tunnistekenttiin lisätään osakenttä \$q, tarkentava tieto
- 502-kentän väitöskirjamerkintä muotoon Diss. + tuplakentät yhdistetään
- Asiasanakenttä 650 ilman sanastokoodia -> 653, kontrolloimaton hakutermi
- Asiasanat järjestetään sanastoittain
- Ajot tehdään joulukuun puolesta välistä tammikuun lopulle
- Myöhemmin sarjakentät, 505, RDA:n edellyttämät fraasikorjaukset

Tuplatietueiden käsittely

- Pasi Tuominen kehittänyt algoritmia entistä tarkempaan tuplien tunnistukseen ja yhdistämiseen – tekeillä tietojenkäsittelytieteen gradu
- Testattu runsaalla Melinda-aineistolla
- Koodia pyydetty myös kirjastojen nähtäväksi, mutta se ei tässä vaiheessa ole mahdollista datan oikeuksien takia
- Varsinaisia muutosajoja ajetaan helmikuusta alkaen

Inhimillinen tekijä vielä tarpeen

- Ohjelmat hakevat ehdotuksia tuplista ja listaavat osan vain kuvailijan katsottavaksi ja päätettäväksi
- Monimutkaiset painosmerkinnät, sarja- vs. monografialuettelointi, kirja vs. nuottijulkaisu, ISBN-tunnus vanhassa kirjassa jne. jne.

Kiitos!

ulla.ikaheimo@helsinki.fi